315

(12) UK Patent Application (19) GB (11) 2 403 342 (13) A

(43) Date of A Publication 29.12.2004

(21) Application No: 0414114.9

(22) Date of Filing: 23.06.2004

(30) Priority Data:
(31) 10603053 (32) 24.06.2003 (33) US

(71) Applicant(s):
Agilent Technologies, Inc.
(Incorporated in USA - Delaware)
395 Page Mill Road, Palo Alto,
California 94306, United States of America

(72) Inventor(s):
Dean R Thompson
William M Old
David Lee Gines

(74) Agent and/or Address for Service:
Williams Powell
Morley House, 26-30 Holborn Viaduct,
LONDON, EC1A 2BP, United Kingdom

(51) INT CL^7:
G01N 30/72 30/86 30/88 , G06F 17/30

(52) UK CL (Edition W):
H1D DMH D12A D12E D21A D21B D21C D51

(56) Documents Cited:
US 5247175 A **US 20040096982 A1**
US 20030109990 A1 **US 20030078739 A1**

(58) Field of Search:
INT CL^7 **G01N, G06F, H01J**
Other: **Online: JAPIO, EPODOC, WPI, TXTE**

(54) Abstract Title: **Method and program for identifying ions from chromatographic mass spectral data sets**

(57) A method and a computer program product for identifying related ions in a input data set produced by analysing a sample. The components in the sample could be peptides. Each row of data in a input data set representing intensities over time for a particular mass to charge range are correlated with every other row of data in the input data set producing a correlation matrix. The correlation matrix is clustered and each group represent covarying chromatograms. At least one time period for each group is selected, and a resultant spectrum for each group is produced. The method and computer program is useful for data sets containing overlapping components, and for reducing the size and complexity of a data set to be analysed.

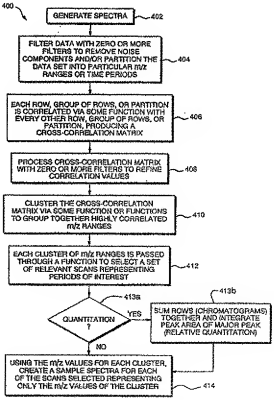
400
GENERATE SPECTRA
402
FILTER DATA WITH ZERO OR MORE FILTERS TO REMOVE NOISE COMPONENTS AND/OR PARTITION THE DATA SET INTO PARTICULAR m/z RANGES OR TIME PERIODS
404
EACH ROW, GROUP OF ROWS, OR PARTITION IS CORRELATED VIA SOME FUNCTION WITH EVERY OTHER ROW, GROUP OF ROWS, OR PARTITION, PRODUCING A CROSS-CORRELATION MATRIX
406
PROCESS CROSS-CORRELATION MATRIX WITH ZERO OR MORE FILTERS TO REFINE CORRELATION VALUES
408
CLUSTER THE CROSS-CORRELATION MATRIX VIA SOME FUNCTION OR FUNCTIONS TO GROUP TOGETHER HIGHLY CORRELATED m/z RANGES
410
EACH CLUSTER OF m/z RANGES IS PASSED THROUGH A FUNCTION TO SELECT A SET OF RELEVANT SCANS REPRESENTING PERIODS OF INTEREST
412
413a
QUANTITATION ?
YES
413b
SUM ROWS (CHROMATOGRAMS) TOGETHER AND INTEGRATE PEAK AREA OF MAJOR PEAK (RELATIVE QUANTITATION)
NO
USING THE m/z VALUES FOR EACH CLUSTER, CREATE A SAMPLE SPECTRA FOR EACH OF THE SCANS SELECTED REPRESENTING ONLY THE m/z VALUES OF THE CLUSTER
414

FIG. 10

GB 2 403 342 A

Method and program for identifying ions from chromatographic mass spectral data sets

Description of **GB2403342** Translate this text

METI-IODS AND DEVICES FOR IDENTIFYING RELATED IONS FROM

CHROMATOGRAPTIIC MASS SPhCTRAI, DATASETS CONTAINING

OVERLAPPING COMPONENTS

RELATED APPLICATIONS

This application is a continuation-in-part of Attorney Docket No. 10020515 -1 (2003309-0034), U.. Patent Application No. 10/388,088, tiled March 13, 2003, entitled " Methods and Devices for Identifying Biopolymers losing Mass Spectroscopy", Dean R. l'hompson and Steven M. lischer, which is incorporated herein by reference in its entirety. 1(,

This application is related to mass spectral analysis, and more particularly to processing mass spectra generated by mass spectral analysis. The invention relates to a method for identifying related ions (or for quantifying at least one ion) in an input data set and to a computer program product for doing the same. i>

Mass spectroscopy is a powerful analytical tool that may be used in identifying unknown compounds as well as their quantities. Mass spectroscopy may also be useful, for example, in clucidatng the structure and chemical properties of molecules, and may be used in connection with organic as well as inorganic substances. 'l'he identification of proteins and other molecules in a complex mixture derived from biological sources may be performed using mass spectroscopy. A variety of different techniques have been developed for use with the identification of molecules, such as proteins.

Prior to performing mass spectroscopy, one technique separates various proteins in the mixture using two-dimensional gel electrophoresis (2DE). The resulting spots may be excised and digested to break the proteins into shorter polypeptidc chains. '11ese digests may be analyzed via mass spectroscopy and the resulting spectrum compared to spectra predicted prom amino acid sequences and information included in databases. The foregoing technique has difficulty, for example, in resolving highly acidic and hydrophobic proteins.

In order to overcome the foregoing difficulties in the first technique, efforts have been made to perform the separation of such mixtures via higl1 performance liquid chromatography (HPLC). rlllese efforts include digesting all of the proteins in the mixture prior to attempting separation techniques resulting in a hyper-complex mixture. I Jsing such a hyper-complex mixture, it may be neither practical nor possible to provide a complete and perfect separation.

Rather, the eluate entering the mass spectrometer may have multiple peptides present at any point in time such that multiple 1loptides coehte resulting in mass spectra that may c,onlain a mixture of ions from the various peptides present.

l'he foregoing may be further complicated by two additional factors. First, large molecules such as peptides may tend to collect a lot of charge during electro-spray ionization.

As a result of the electro-spray ionization and the collection of a large charge, the spectrum of each pepticlc may have multiple peaks corresponding to the multiple charge states.

Additionally, high-resolution mass spectrometers, such as the time of flight devices, may resolve multiple isotope peaks for each charge state As a result of the above factors, a very complex spectrum may result.

In order to reduce the complexity of the resulting spectra, techniques, such as charge assignment and de-isotoping, may be performed. However, these techniques may be sensitive to various types of interference and noise, chemical as well as electrical.

S Additionally, a complete data set of spectra produced by, for example, liquid chromatography/mass spectron1etry processing (LC/MS) may be quite large. A spectrum may be taken at various frequencies, such as several times a second or every few seconds, over a period of several hours. The size of such a data set presents a number of challenges in accordance with analyzing such a large amount of data.

One technique to reduce tile computational burden in connection with such large amounts of data is to only select particular spectra to be analyzed in detail in accordance with particular criteria. llowever, these spectra are typically selected manually by visual inspection of the chromatograpl1ic data, which may lie time consuming, choosy, and error 1 5 prone Accordingly, it may be desirable to provide a technique for analyzing chromatographic information, such as may be included in an L. C/MS dataset, and using the resulting analysts infonnation to separate related ions into spectra representing individual compouncls. It may also be desirable to use tl1e resulting analysis information to identify the particular spectra that provide maximum signal levels for subsequent analysis. It may also be desirable to remove and filter noise from the data and significantly reduce the size and complexity of the dataset to be analyzed. It may also be desirable to use such a technique in connection with protein identification as well as be generally applicable for the analysis of other classes of molecules sharing similar characteristics.

In accordance with one aspect of the invention is a method for identifying related ions in an input data set produced by analyzing a sample comprising: correlating each row of data in an input data set with every other row of data in said input data set producing a correlation matrix, each row representing intensities over time for a particular mass to charge (m/z) range, each element of said correlation matrix including a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said correlation value; clustering said correlation matrix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each group representing covarying chromatograms; selecting at least one time period of interest for each group; and producing a resultant spectrum for each group by sampling chromatograms included in each of said groups at each of said at least one time period of interest of using a form of said input data set.

In accordance with another aspect of the Invention is a method for quantifying at least one ion In an input data set produced by analyzing a sample comprising: correlating each row of data in an mput data set with every other row of data in said input data set producing a correlation matrix, each row representing intensitles over time for a particular mass to charge (m/z) range, each element of said correlation matrix including a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said correlation value; clustering said correlation matrix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each group representing chemically related components exhibiting correlated chromatographic behavior; selecting at least one time period of interest for each group; and producing a resultant spectrum for each group by samphng chromatograms included in each of said groups at each of said at least one time period of interest of using a form of said input data set.

In accordance with another aspect of the invention is a computer program product for identifying related ions in an input data set produced by analyzing a sample comprising: machine executable code that correlates each row of data in an Input data set with every other row of data in said input data set producing a correlation matrix, each row representing Intensities over time for a particular mass to charge (m/z) range, each element of said correlation matrix including a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said correlation value, machine executable code that clusters said Correlation matrix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each group representing eovarying chromatograms; machine executable code that selects at least one tmle period of interest for each group, and notching executable code that produces a resultant spectrum for each group by sampling chron1atograms included in each of said groups at each of said at least one time period of interest of using a form of said input data set In accordance with yet another aspect of the invention is a computer program product for quantifying at least one ion in an input data set produced by analyzing a sample comprising: machine executable code that correlates each row of data in an input data set with every other row of data in said input: dale set producing a correlation matrix, each row representing intensities over time for a particular mass to charge (m/z) range, each element of said correlation matrix inchding a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said correlation value; machine executable code that clusters said correlation matrix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each group representing chemically related components exhibiting correlated chromatographic behavior; machine executable code that selects at least one time period of interest for each group; and machine executable code that produces a resultant spectrum for each group by sampling chromatograms included in each of said groups at each of said at least one time period of interest of using a form of said input data set.

Preferred features and advantages of the present invention will become more apparent from the following detailed description of exemplary embodiments thereof taken in conjunction with the accompanying drawings in which: Figure I is an example of a block diagram illustrating processing steps of a substance input to a mass spectrometer; Figure 2 is an example of an embodiment of a computer system included m Figure 1; Figure 3 is an example of an embodiment of a host included in the computer system of Figure 2; Figure /1 is an example of a limctional block diagram of components included in a mass spectrometer of Figure 1, Figures 5-9 are example graphical illustrations of alternate displays of data output from the mass spectrometer of Figure 4; Iiigure I O is a flowchart of method steps of an example embodiment for performing Ion identification and filter processing upon data output from the mass spectrometer of Figure Figure 11 is a flowchart of method steps of an example embodiment for processing different types of mass spectral data sets; Figure 12 is a flowchart of method steps of an example embodiment for performing clustering or grouping of highly correlated rows as used In Figure 12 flowchart processing steps; and Figures 13-17 are example graphical illustrations of data sets at various processing steps of the method of Figure 11.

Referring now to Figure I, shown is an example of a block diagram of processing steps that may be performed in connection with identification of a molecule within a mixture in an embodiment. In this particular example, the substance may be a mixture of one or more molecules, for example, such as peptides or proteins, being processed for identification. It should be noted that the techniques described herein may also be used in performing a quantitative analysis of molecules in a sample. An input sample or substance 12 is digested m the enzymatic digestion processing 14. This enzymatic digestion processing 14 breaks the proteins in the sample] 2 into shorter polypeptide chains. Subsequently, the digests may then be separated via a separation processing technique 16. Any one of a variety of different separation processing techniques may be used SUCH as liquid chromatography, 2D Gel separation, and the like it should be noted that generally any separation technique and/or digestion technique may be used to separate the various polypeptides in accordance with, for example, molecular weiglit, electrical fields and the like After separation processing 16, the resulting separations may be Input to a mass spectrometer 18 producing mass spectra data 2() as an output. The mass spectra data may be input to ion identification and filter processing 24. 'Ihe Ion identification and filter processing 24 may use a computer system 26 in comlection with performing processing steps therein Details about the specific processing steps performed in connection with the ion identification and filter processing 24 are described elsewhere herein in more detail Subsequently, output of the ion identification and filtering processing 24 may serve as an input to post-processing 22.

Post-processing 22 may Include, for example, performing de-isotoping or charge assignment. Post-processing 22 may also Include for example, comparison of monitored output data to known spectral data, for example, In order to identify a particular known type and quantity associated with proteins and the like that may be included In the sample 12. 'The post-processing 22 may also use the computer system 26. It should be noted that post- processing 22 may use the same or different computer system used in connection with the processing steps of the ion identification and filter processing 24. As an output of post processing, sample information results 23 may be produced. The results 23 may include, for example, types of known proteins and quantities identified in the sample I 2.

It should be noted that, although the particular sample or substance 12 described in the foregoing and throughout this example may be a protein, the techniques described herein may be used in connection with other types of substances or samples 12 to Identify other molecules and/or associated q-mtties. An embodiment may include adtltional and different I 5 processing steps than those described herein in accordance with the type of sample or substance 12 being analyzed as well as the particular components being identified within the sample or substance. This may affect the processing steps performed both before and after processing by the mass spectrometer Iior example, the enzymatic digestion processing may not be used in connection with performing an analysis of a sample or substance that does not include proteins.

Referring now to Figure 2, shown is a more detailed example of an embodiment of the computer system 26. It should be notetl that Figure 2 illustrates only one particular arrangement ol'a computer system that may be included In the embodiment 10 of Figure I. The computer system 26 mchdes a data storage system 112 connected to host systems 114a- 1 14n, and a data manager system] 1 6 through communication medium 118 In this embodiment of the computer system 26, the N hosts 1 1 4a-1 1 4n and the data manager system 116 may access the data storage system 1 12, for example, in performing input/output (I/O) operations or data requests. rl he communication medium 118 may be any one of a variety of networks or other type of communication connections as known to those skilled in the art. The communication medium 118 may be a network connection, bus, and/or other type of data link, SUCH as a hardwire or other connections known in the art. For example, the communication medium 118 may ire the Internet, an intranet, network or other connection(s)] O by which the host systems 1 1 4a- 1 1 4n, and the data manager system may access and communicate with the data storage system 112, and may also communicate with others inchded In the computer system 26 leach of the host systems 1 1 4a-] 14n, the data manager system 1] 6, arid lhe data storage system I 12 included In the computer system 26 may be connected to the communication medium 1 18 by any one of a variety of connections as may be provided and supported in accordance with the type of communication medium 118. The processors included in the host computer systems 114a-1 14n and the data manager system 116 may be any one of a variety of commercially available single or multi-processor system, such as an Intel-based processor, IBM mamtrame or otl1er type ol commercially available processor able to support incoming traffic in accordance with each particular embodiment and application.

It should be noted that the particulars of the hardware and software included in each of the host systems 114a-1 14n and the data manager system 116, as well as those components that may be included in the data storage system I I 2 may vary with each particular] 1 embodiment. Each of the host computers 114a-1 14n, as well as the data manager system I] 6, may all be located at the same physical site, or, alternatively, may also be located in different physical locations Examples of the communication medium that may be used to provide the different types of connections between the host computer systems, the data manager system, and the data storage system of the computer system 26 may use a variety of different communication protocols such as SCSI, F!S(:ON, lilbre Channel, or GIGE (Gigabit Ethemet), and the like. Some or all of the connections by which the hosts, data manager system 116 and data storage system 1 12 may be connected to the communication medium I] 8 may pass through other communication devices, such as a Connectrix or other switching equipment that may exist such as a phone line, a repeater, a multiplexer or even a satellite.

Each of the host computer systems as well as the data manager system may perform different types of data operations in accordance with different types of administrative tasks.

in the embodiment of ldgurc 2, any one of iLc host computers I 14a-l 14n nosy i.ssre a data i 5 request to the data storage system 1 12 to perform a data operate For example, an application may be invoked in connection with ion identification and filter processing 24 and may execute on one of the host computers 114a-1 14n.

It should be noted that the computer system 26 included in the system 10 of Figure I 2() may also be a single computer, such as a personal computer, as well as another arrangement of a plurality of computer systems as described above.

Referring now to Figure 3, shown is a more detailed example of an embodiment of a host computer system I 1 4a- 1 1 4n that may included in the computer system 26 11le host computer system 1 14a may include components such as one or more processors 130, a memory 132, one or more data storage units 134, as well as a display 136, and one or more input devices 138. All of these components within a computer system 1 14a may communicate and transfer user data and command information using a local bus 140.

It should be noted that the components included for the host computer system 114a may also be those components included in an embodiment in which the computer system 26 is a smgle computer, for example, such as a single personal computer that may be used in connection with postprocessing and ion identification and fi lter processing 24.

Referring noNv to Figure 4, shown is an example of an embodiment of a mass spectrometer 18. A mass spectrometer may be characterized as an instrument that measures the mass to charge ratios of individual molecules that have been converted into ions. As described in the following paragraphs, a mass spectrometer does not actually measure the moleculamnass dircct.]y, but rather determncs the mass-lo-charge ratio of the ions formed from a particular molecule or molecules A usel'ul unit for purposes described herein is a unit ret'erring to a fundamental unit of charge, the magnitude of the charge on an electron. T1lG charge of an ion may be denoted by the integer number z of the fundamental unit of charge and the mass-to-charge r atio may be referred to as m/z Figure 4 includes the different functional units of a mass spectrometer that may be represented conceptually in the block diagram 18 of Figure 4 A sample may be introduced via an inlet 156 into a vacuum chawher. it shonicl be noted that a sample may be m any one of a variety of different forms including, for example, a liquid solution, embedded in a solid matrix, or a vapor. Depending on the type of inlet and ionization techniques used, the sample may already exist as ions in solution, or it may be ionized m conjunction with its volatilization or by other methods in the ion source] 50. In this embodiment, as the sample IS introduced into the inlet 156, the sample is placed in a gas phase and then charged to produce ions. The lons are sorted by an analyzer 152 according to their mass-to-charge or m/z ratios and then collected by an Ion detector 154. In the ion detector I 54, the ion flux may be converted to a proportionate electrical current. Output of the ion detector 54 serves as an mput to the data system I 58 recording the magnitude of the various electrical signals as a function of the m/z ratios and converting the information into mass spectrometer data 20.

It should be noted that in the foregoing general description regarding a mass spectrometer, dit'ercnt types of mass spectrometers may vary from the components included in Figure 4 For example, the ion sorting described above may be included in a quadrupole instrument but not in a TOl; mass spectrometer since the TOl; mass spectrometer measures the flight time of the ions in a fixed length tribe. The techniques described herein may be Scud with any type of mass spcctromctcr and any description to a particular type ol' mess I 5 spectrometer should not be construed so as to limit the application of the techniques described herein.

It should be noted that an embodiment may include ion selection processing as part of ion sorting 152 in which only a portion of the particular ions are selected for further 2() processing and analysis. As will he shown and described elsewhere herein, the mass spectrum data output from the mass spectrometer 18 is generally a graph of ion intensity On the y axis as a function of the mass-to-charge ratio (m/z) he displayed on the x axis of the spectrum. It should be noted that the ions coming from the mass spectrometer 18 may be positively as well as negatively charged.

As described herein, the sample may be in any one of a variety of forms when introduced into the inlet 156. For example, if the sample is a solid, the sample may be evaporated or sublimed into a gas phase such as, for example, by heating. Gases and liquids may be introduced through inlet designs which control the flow. Some embodiments may combine various techniques in processing, for example, such as volitization and ionization occurring at the same time. The sample may also be a mixture in which the individual components may be separated prior to input and analysis by the mass spectrometer.

Separation is described In connection with processing step 16 of Figure 1. Separation may be used to simplify mass spectra for a sample with multiple components by reducing the] O number of co-ehting compounds. Gas chromatography may be coupled with mass spectrometry as a means for separation as also described herein. Gas chromatography for example may allow compounds already in a vapor phase to enter the mass spectrometer separated in time so that components of mixtures may be detected and analyzed Liquid chromatographs may also be used as well as capillary electrophoresis devices and other types 1 S of hardware and/or software used in comechon with performing the separation processing prior to introduction of a sample into a mass spectrometer I 8.

Molecular and fragment ions may be produced in the ion source 150 as shown in Figure 4. If the mput is not already ionized, any one of a variety of different ionization 2() techmques may be used, lor example, including elcctro-spray ionization (ISSI). It should tic noted that although both positive and negative ions may be generated m the ion source at the same time, a single polarity may be recorded at any particular time A given mass spectrum may include positive or negative ions. The ions are then input to the ion sorting or analyzer 152. We analyzer may use dispersion or filtering to sort ions according the mass-to-charge ratios or other relative properties Analyzers may include for example magnetic sectors, quadrupole mass filters, Fourier transform ion cyclotron resonance spectrometers, time of flight mass analyzers and the like. Subsequently, the sorted ions produced by the ion sorter or analyzer 152 are input into the ion detection processing] 54 where the particular charge of the ions are determined. s

It should be noted that a computer may be used in connection with controlling the mass spectrometer as well as in spectrum acquisition, storage and presentation. dais described herein for example in connection with the processing of the block diagram I O of Figure 1, software and/or hardware may be used in a computer system in connection with performing quantizatori, spectral interpretation, and compound identification It should he noted that in addition to the 1SSI technique to generate ions as a result of the source processing I 5() within the mass spectrometer, chemical ionization, Resorption ionization, elecro spray ionization, and the like may he used in comect.ion with performing I S ionization. It should lie noted that. fair polypepticles, and the. like (biomoleculc s), techniques such as ELI, Matrix Assisted I user Resorption ionization (MAI. OI), Atmospheric-Pressure MALDI (AI,-MAI,T)I), and other "soft" ionization techniques are preferred over "hard" ionization techniques. Soft and hard with respect to ionization techniques refer to the energy levels used to ionize the molecules of interest. flard ionization techniques are not compatible with liomolccrles l:'ecacse they result in extensive fragmentation.

Separation techniques, such as gas chromatography (GC), liquid chromatography(T.(), and tile like as described herein may be used in connection with mass spectrometry in order to identify chemical compounds. In connection with using a mass spectrometer (MS) with a gas or liquid chromatograph, an interface may be used to restrict or reduce the gas flow into the mass spectrometer For example, this may result in an interface being introduced in between separation processing 16 and mass spectrometer I 8 as shown in connection with Figure 1. ATIY chromatographic technique, SUCh as, for example, LC, C, FE7Electrophoresis, and the like may be used in connection with biomolecules. The use of liquid phase techniques may be preferred due to the ease with which they may be interfaced with a mass spectrometer in addition to the ability to monitor the chT-omatogTaphic behavior of elating components.

In conTlecton with GC/MS, I.C/MS or other combinations, the output data of the mass spectra 20 consists of a series of mass spectra acquired over time. To generate this information, the mass spectrometer may scan the mass range, for example, for a particular m/z range repeatedly for a particular chromatographic run. scan may be taken at a predetermined frequency, such as, for example, every second, or several times a second.

The particular scats I'rclueTlcy selected may vary in accordance with an embodTTTTent An embodiment may select a scan frequency that varies with the average expected peak width and may be, for example, an order of magnitude greater than this. In one embodiment, the mass spcct:rometer scans at a rate which is 1 O-fold higher than the rate at which compounds are elating. This translates to at least IO scans over an average chromatographic 2() peak.

Referring now to ldgure 5, shown is one form of a graphical representation of the spectral data as may he displayed. (graphical display 2()0 of Figure 5 shows a total jOTT chromatogram (TIC). The TIC represents the intensities of all the ions as summed in l7 connection with each particular scan. Thus, the TIC represents an aggregate amount of ion intensity in each scan.

Referring now to lTigure 6, shown is an example of a graphical representation 250 of how a TIC 260 may further be represented by a plurality of individual ion profiles 270. A particular point 271 a in the TIC 260 may be represented by summing the individual ion profiles 271b as illustrated in 270 along the direction indicated by arrow 272. Figure 6shows alternative data displays of chromatographic data as may be output from the mass spectrometer 18.

It should be noted that in connection with capturing spectra at a particular frequency, the particular frequency may vary in according with each embodiment. For example, with techniques described herein, spectra may be gathered several times every second It should lee noted that 'I'I(s are clfect.cd by noise c-'ml':,nents of the data set Referring now to Figure 7, shown is an example of another form of how data output from a mass spectrometer may be displayed. 'I'he data display 280 may be referred to as a contour plot where the scan number is on the x axis. The particular m/z value is represented on the y axis with the intensity represented as a gray scale value Viewing a slice vertically through the representation 280 of Figure 7 results In a spectrum for a particular clution time.

A horizontal slice of the graphical illustration 280 of Figure 7 represents the ion current for a particular m/z value over time which is commonly referred to as the extracted ion chromatogram (XIC) Referring now to Figure 8, shown is an example of the graphical representation 300 of an XTC. The Illustration 300 represents an XIC for an m/z ratio of 100 over time.

In connection with the XICs, it may be noted that two or more components of an original mixture may co-elute at a particular point in time. llowever, the elusion profiles of each of the respective two components in most cases will exhibit differences over a series of time points or scans. It should also be noted that ions resulting from the processes of the mass spectrometer may tend to co-vary chromatographically by exhibiting similar elusion profiles Referring now to Figure 8, shown is an example of a graphical illustration 350 representing XICs for four different m/z values overlaid. All four m/z values are co-eluting at a scan point I as identified on tile illustration 350. llowever, note that only ions 3 and 4 are. co-var),ing. (2o-varymg ions in this example may he visible in a contour plot as shown in 1 igure 7 as a sere.s of horizontal bars arranged in a column. I lowever, when the XI(..s of the corresponding ions 3 and 4 are examined, similarity n1 elusion profiles may be observed These observations regarding covariance may be utilized in the processing steps described herein.

Referring now to Figure 10, shown Is a flowchart of processing steps that may be included m an embodiment of the ion identification and filter processing 24 previously rlescriber1 in connection with ligure 1. At step 402, tile spectra are generated as a result of mass spectrometer processing, for example, an T.(/M,S data set of a time series of spectra Tile data set may be represented as three columns of data including a scan number, an m/z value, and a corresponding intensity. 'I'his may be as represented In example display 280 of Figure 7. The same set of input data may also be represented as one or more XICs described elsewhere herein ITI whicl1 each m/z value is monitored over time. Each XIC is the scan number or time on the x axis with the intensity monitored over time on the y-axis. There is an XIC for each m/z value. The format of the data used in connection with the processing steps described herein is a twodimensional matrix having a row index on the Y axis of the m/z ratio, and a Cohen index on the axis of a scan number. The value within a cell or entry identified by a row and column is the associated intensity value At step 404, the data may be filtered with zero or more filters to remove noise components and/or partition the data set into particular m/z ranges or time periods. It should be noted that in order to reduce the "noise" in the data set being analyzed, the choice of Titers anti the particular cornbinaticn and order used may vary depending on the quality of the data For example, m one embodiment, the following filtering techniques may be used: I truncate data below a ccrtam thT-esholci 2. median filter 3. 2-D gaussian convolution filter 4. remove 1)C noise using 1)(: filtering techniques 2() These and other filtering techniques may be founci, for example, in Pratt, W.K, entitled "I)igtal Image Processing", by John Wiley & Sons, 1991, New York.

IJsing the foregoing types of filecring techniques in one example embodiTnent, the output of the filtering processing of step 404 is a data matrix with the same number of columns (scans or time points) as the original matrix. An embodiment may have a reduced number of rows as a result of step 404 processing In comparison to the number of rows in the original data set due to removal of the zero rows generated by filtering of noise. The magnitude of the data reduction depends on the cutoff threshold in step I above, as well as other filter parameters used in connection with steps 2-4 processing that may be utilized in an embodiment. In one embodiment in connection with steps 1-4 as outlined above, the foregoing parameters may be used with associated processing steps: step 1) Inmcate values less than 5 X, of maximum, step 2) 5xS median filter, and step 3) use a Gaussian filter with a width approximately that of the expected width of the chromatographic peaks. In connection with flt:ering step 4 denoted above, no parameter selection is necessary. It should be noted that the foregoing techniques, as well as guidelines for their use, are well known.

An embodiment may use any combination of hardware and/or software to implement the l'oregoing filtering processing in an embodiment. In an embodiment using software to Implement the foregoing filtering steps and other processing dcscrilled herem, any one or more programming languages, such as, for cxampic, (I, (a i, Java, I;O'I'KAN, and/or any] 5 one or more software packages, such as, for example, MA'1'I AID, may be used. 'I'he particular ones may vary In accordance with what is available in each implementation As an alternative, or in addition, to filter processing at.st:ep 4()4, an embodiment may partition tlc data set k' reduce tle number of rows in the data matrix One embodiment may select only those rows of data within a particular m/z range I;or example, data peaks may be determined and a particular m/z range may be selected for a range of values on spanning a data peak. Use of partitioning in this processing step refers to a process of data reduction. At some point, partitioning may become necessary m an embodiment because of memory constraints clue to the size of the resultant correlation matrix formed and used in other processing steps described elsewhere herein. 'I'he size of the correlation matrix depends on the number of rows in the original data matrix (number of non-zero mass samples).

Consider, for example, an embodiment performing the processing steps described herein in connection with flowchart 400 using time of flight ('I'OF) datasets having greater than 100,000 mass samples for each spectrum in the dataset. If all m/z rows of the data set are considered, assuming that there Is no tamcation or filtering, then the correlation matrix has lelO elements, which at 4 bytes an element, results in a 39 GB matrix. An embodiment may utilize the}partitioning technique to reduce the size of the matrix.

Refen-ing back to I;igure 7, graph 280 may be represented by a data set in matrix 1 () fond, for example, having approximately 250 m/z rows in the dataset represented. Actual datasets tend to be much larger, but this serves as a good example. Referring to the graph 280, G major peaks may be discerned A peak finding routine may be utilized to locate the mayor peaks with reference to a particular scan number. One peak fmding technique that may lc used m an embodmcut ls based on the calculation of denvativcs. For example, at the peak maxnnrm, the first derivative is /ero and the second derivative is negative The peak finding routine may be performed ha the time and m/z dimension to find the peaks. A range of scans may be selected, peak +/- range value, as well as examining only scans for the maxima 'I'he multiple rows ha each peak may be reduced by, for example, combining the rows by adding them An embodiment may also take the median of samples. An embodmcnt may also select the maximum representative row for the mass peak. Another embodiment may inclrdc the use of Image processing algorithms, such as the watershed algorithm, to perform peak finding in the time arid m/z dhnonsions simultaneously. 'I'he watershed algorithm, as well as other image processing techniques are known in the art and described, for example, in K.R C:astleman, "Digital Image Processing" Prentice-I-latl Inc., New Jersey 1996. In this embodiment using the watershed algorithm, the dataset is treated as an image, for example, as shown in Figure 13. First, the local maxima are determined using an extended-minima transfonn and imposed on the image as described, for example, in Pierre Soille, Morphological Image Analysis: Principles and Applications, Springer-Verlag, 1999, pp. 17()-171. This helps reduce oversegmentation during subsequent processing steps. Next, watershed segmentation is performed on the image which detects the peak boundaries (in time and mass) and segments peaks which are not fully resolved. IJsing the foregoing has several advantages. The peaks, which consist of multiple mass rows or chromatograms, may be combined into a single peak chromatogram by summing all of the intensities within the peak boundary m a row-wise manner. I he peak chromatograms may then serve as an inputs] O to the grouping algorithm, rather than using every mass row in the dataset. This results in a significant reduction in the number of rows input to the grouping algorithm, and a smaller size of the resultant correlation matrix Additionally, peak splitting is no longer neccessary with this technique, smce the peak detection performs this automatically Furthermore, quanttation may be pericned by sun mug the intensifies within the peak boundaries Using any one of the foregoing results in collapsing the multiple rows into one peak.

It should be noted that different techniques used here may effect subsequent processing steps.

For example, if rows are added together, the processing at step 414 in figure 10 is also affected. Without such peak fmding routines, multiple rows of data are used for a single 2() peak in a data matrix as input into a correlation routine, which is redundant due to the hilly correlation of rows within a single peak. Referring back to the example dataset with 250 rows, this may be reduced to a matrix of 6-10 rows, corresponding to the number of peaks, and reduces the size of the correlation matrix as well.

It should be noted that the partitioning may be preferred to filtering for a large data set, for example, greater than 10,000 m/z samples, due to the computer resources and time required for performing processing ol the large data sets.

At step 406, each row, group of rows, or partition is correlated using some function with every other row, group of rows, or partition producing a corTclation matrix representing the degree to which the rows are related to one another. ISach row represents intensities over time for a particular m/z range. The resulting correlation matrix is a two dimensional matrix symmetrical about the diagonal such that the diagonal entries are 1 and the upper and lower I O triangular portions are Identical. In other words, each entry having indices ",j" Is the same value in the entry having Indices "j,i" . The correlation for two rows x and y may be represented as: n Ti -- Ant * Eli -- II?} I --- .= 2 31 ( ) 2 in which "mx" represents the mean Tahoe of row x, "my" represents the mean value of row y, and the index "i" ranging from 1 to n represents the index of the entry In the row with n being the total number of rows At step 40X, the correlation matrix is processed with zero or more tillers lo further refine the correlation values. At step 410, the cross correlation matrix may be clustered using some function or l; nctions to group together highly correlated m/z ranges or identify clusters of n/z ranges. One particular clustering or grouping technique is described elsewhere herein in more detail. fin embodiment may also utilize other clustering or grouping techniques such as, for example, hierarchical clustering, K-means clustering and others. Such techniques are described, for example, in Seber, G A 17., Multivariate Observations, Wiley, New York, 1984, and Spath, Il., Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples, translated by J. (oldschmidt, I-Ta] sted Press, New York, 1985.

At step 412, each cluster or group of m/z ranges is passed through a function to select a set of relevant scans representing periods of interest. In one embodimeT1t, the one or more scans may be determined by first determining a maximum point by summing the intensities of the XICs at each scan point within each group, for example, by adding the rows of the data set for all TOWS withm each group. 'I'he scan corresponding to the maximum point or peak intensity may be determined as a scan of Interest. An embodiment may also determine more than one scan of interest by determining a scan range, for example, utilizing the peak or maxiTIl'm value. The scans of interest selected may he those scans falling within peak+/ range value, where range value may vary with an embodiment. 'I'he range value may be, for cxamplc, / the peak vahTe 1 r One technique for selecting the range of a chromatographic peaks is to select the range that TS full width at half maximum (EWI-IM), meaning that one selects the range between the two poiTTtS On either side of the peak that are at half the height of the peak. Other embodiments may rise other techniques for range deterTnination As described herem, the scan(s) of interest may vary with embodiment. An embodiment inay determine a single poinl as a scan of hteresl representing, for example, the maximum average ion signal for the selected m/z values or the time centroid of the cluster.

ATl embodiment may select a range of scans, stick as the complete set of scans containing a signal for selected m/z valises, and the like. More than one scan may be selected, for example, If the signal is weak and/or there is excessive noise to increase the signal to noise ratio. One technique sums all columns containing a signal for the group to maximize the signal.

Control proceeds to step 41 3a where a determination Is made as to whether quantitation is being performed. Quantitation generally refers to the processing step of determining an amount or quantity of molecule rather than identifying a particular type or types of molecules If quantitation is being perfonned, control proceeds to step 41 3b where rows (chromatograms) are added together. Relative quantitation is performed by integration of a chromatographic peak to obtain the peak area, which is proportional to the quantity of the component in the mixture. The foregoing integration sums the intensities for a given m/z range between two time points spanning the peak of Interest At step 414, the m/z value(s) for each cluster fir group as molded in the input data set may be used to create a sampled spcctnm for each ofthe scans selected In step 412 representing only the m/z values ot'tlle cluster. In other words, for each one or more scan vahes of interest, a corresponding column of intensities fiom the original data set is used to produce a spectrum nor each group. It should be noted that when performing step 414 processing, an emboLliment may utilize the original data set or a filtered form of the original 2() data set to produce the resulting spectra.

Tllc input data produced at step 402 used in the foregoing processing may be gathered by running the mass spectrometer at nonnal energy levels (U spectrum), high fragmentation energy levels (1. spectrum), or in alternating scan mode producing alternating U and It spectra. Wlen using alternating scan mode producing data sets including alternating U and 1 spectra, the chromatographic correlation of the parent peptides (U spectra) and their respective fragment ions (F spectra) may be used to associate parents with their fragments.

This characteristic of time or scan correlation between parents and associated fragments may be used, for example, in cases where multiple parents are being fragmented simultaneously, but exhibit sufficient differences in their respective elusion profiles. The respective differences in the elusion profile enable differentiation between the dil'ferent parents to be matched with appropriate fragments.

If the input data is produced using the alternating scan mode, two different approaches l O may he used in processing the input data. Both approaches are described in following paragraphs. In a first approach, the lJ and F spectra may be combined In a second alternate approach, the U and F spectra may be processed separately.

l;orthe first approach, tlie 1J and corresponding 1; spectral pairs are added together 1 s prior to perfinMing step 40G. It should be noted that the 1 spectrum may be filtered prior to performing the summation of the F and corresponding U spectrum. 'l'hs filtering may be performed, tor example, clue to the lower intensity of fragmentation spectra. In one embodiment, a combination of baseline sullb-action, Kalman smoothing and Savitzky- (Jolay filtering are performed. Subseciuent to performing the summation, acliditonal filtering may also be performed on the composite spectra. Correlation, filtering, clustering, selection of relevant scans and other processing associated with steps 406, 408, 410, and 412 then proceed as described c]sewherc herein resulthg in a set of component spectra (U and combined) In following paragraphs, this may be referred to as the A set. When performing processing associated with step 414, two different spectra are created - one from the original U spectrum at a selected scan for a group, and a second 1' spectrum sampled at the same scan.

In the first approach, the precursor (parent) ions may be identified by first deriving the A set spectra representing the combined U and 1;, and then sampling the original U-only dataset at the masses present in set A, and at the scan maximum identified for set A. '11,e parent ions are where there are intensities at the sampled masses iT1 the lJ-only spectra.

The combined spectra in the A set, assuming that no parents have exactly the same chromatographic profiles, should contain the parents m/Z value with fragments from only that parent. lathe next step TS to determine which m/z value in this A spectrum is the parent.

The m/z values identified ITI the A spectrum are then used to sample the original IJ spectra at the scan maxiTnum identified for spectrum A. Intensities occurring at these sampled masses in the U spectrum indicate the parent iOT1 masses. Absence of signal at a sampled m/z indicates a fragment ion. By performing the foregoing, the parent masses are identified within the c.onhned [I-ii c-mponcnt spectrum, spectnrn In addition to the first summation approach, a second time correlation approach may be utilzcd. Correlation processing of step 406 may be performed OTI the U and F datasets separately. The IJ and F spectra may be sampled at the scan values as described above in alternating mode It should be noted that to utilize this second approach, the 1. spectra should 2() have a sufficient signal to noTsc ratio for satisfactory correlation. If this is not the case, the summation technique may perform better. Additionally, as with the summation method, filtering techniques may be performed on each of the F and/or U spectra. It should be noted that different filtering techniques may be utilized in an embodiment on the F spectra due to the typical lower signal to noise ratio making the F spectra more error sensitive. As in the summation method, there should be a 1-1 correspondence between the spectra in both the U and F sets, the parents in the sets from the U. and the fragments in the sets from F. correlated in time.

Referring now to Figure 11, shown is a flowchart 600 of method steps of one embodiment for perfonning processing of input spectra produced using a mass spectrometer operating in alternating seen mocie. I7lowehart 600 summarizes the processing steps described above.

At step 602, a determination is made as to whether the input data set includes alternating U and F spectra. If not, control proceeds to step 604 where the processing steps described in connection with flowehaTt 400 may be perfonned to process the input data set.

Otherwise, control proceeds to step 606 where determination is made as to whether any filtering is performed upon the separate U and/or F spectra If so, control proceeds to step 6()8 where flee filtering is perlirmed prior to step (i I (). At step 61 O. a determination is node as 1 lo whether the surnmatTon teehniclue, the first approach described above, is to be performed.

If so, control proceeds to step 616 where IJ and adjacent F spectra are added together. At step 618, filtering may be optionally performed on the combined U-F spectra. At step 620, the correlation and otheT processing steps, such as 406, 408, 410, 412 and 414 described in flowchart 400, are performed producing a resultant combined T J-F spectra referred to as set A. At step 622, the m/z values Identified in the A spectrum are then used to sample the original U spectra at the scan maximum identified for the spectrum in set A. At step 624, parent ion m/z values are detenninecl to be those having an intensity value > (). Absence of a signal at a sampled m/z valiTe such that the intensity = 0, indicates a fragment ion.

If at step 610 it is determined that the summation technique is not used, the alternative second approach, the time correlation approach, is utilized At step 612, correlation and other processing steps, such as 406, 408, 410, 4] 2 and 414 described in flowchart 400, are performed separately on the tJ and F spectra. At step 614, the parents are matched to corresponding fragments utilizing the correlation of time centroids for the processed U and F gTOtlp S. It should be noted that the mass spectrometer in alternating scan mode may utilize a scan rate that is much higher than the rate at which components are elating. For example, in one embodiment, the scanning rate is a factor of 10 or more than the rate at which components are elating from the mass spectrometer. Selected scanning rates are described elsewhere herein.

I f the input data set includes only tJ spectra with no l'ragmcnts, the analysis is perl'onned to examine each peptide in the mixture, or molecule in the sample. Each group corresponds to the charge states and isotopes of a smgle peptide or molecule coeluting at the same time When the input data set includes only IJ spectra, the techniqtes described herein may be used to determine which m/z ratios of peaks are of the same peptide or molecule.

l'his may be a useful preprocessing step prior to performing, for example, charge assignment, isotope clustering, de Provo sequencing, database searching, and the like.

If the input data set inchdcs only 1 spectra, each group corresponds to the charge states, isotopes, and fragments of a smgle peptide or mo]ecu]c coehting at the same time.

Refenring now to Figure] 2, shown is a flowchart 700 of method steps of an example embodiment of' a clustering or grouping process. The method steps of flowchart 700 may be performed as part of step 410 processing The input at step 70? is the correlation matrix, C, produced as a result of step 406 processing. At step 702, the row "i" of the matrix C is determined as the row with the largest magnitude. The magnitude of a vector may be dehmed in dif.'fercnt ways. For example, in one embodiment, the magnitude may be defined as a p-norm of a vector for l <= p <= infinity, p being an integer value, for a vector x as: ! 11 X llp= [1 xj lP]P /=! The vector x may include "n" values that are each real or complex elements. In the instance where p = infinity, the following is true: ¦¦ x ¦¦,,,= ma,x ¦ x ¦ An embodiment may also use other types of norms in delenninmg a magnitude, such as, for example, other norms involving derivatives, such as the Sobelev none. Other measures of magnitude that may be included in an embodiment inchde. a number ol'elements above a thrcsho]d, entropy, concentration, logarithm of energy, and the hke as descried in, for 2 () example, Wickhauser, "Adapted Wavelet Analysis from'l-'heoy to Software", 1994, A.K Peters, Massachuetts, and Atkinson, "An Introduction to Numerical Analysis", 1989, John Wiley and Sons, USA.

At step 704, a determination is made as to whether the magnilrde is less than a first threshold, or if all rows have been processed. If either condition is tree, processing stops Otherwise, control proceeds to step 70G where a new group is started with the selected row "i" included in the new group. Scan "S" at which row "i" maximizes is also determined and used as a criteria for grouping subsequent rows. The first threshold may vary with each embodiment and may be empirically determined m accordance with each particular data set and mass spectrometer settings and characteristics. lior example, in one embodiment the first threshold may be.15 specifying a minlmum correlation value. If this Most threshold is increased, the number of groups may decrease. fLt step 708, a counter ";" is initialized to be the value of "it-l" At step 710, a determination is made as to whether the current element, Cij is greater than a second threshold, and whether the peak of row ";" is within a certain number of scans (threshold 3) of scan "S" (peak scan for row "i"). For example, in one embodiment, this second threshold may be.75 and the third threshold = 2 scans. [f CiJ is greater than the threshold 2, and the scan difference is less than threshold 3, control proceeds to step 712 where row j is added to the current group if the row j has not already been considered. At step 714, rOWJ is excluded from further consideration and conho1 proceeds to step 716. Ii, at slep 710, it is determined that (I'd is not greater than Lhe.seconcl threshold, control proceeds direc fly to step 71 (i It shooed be noted that the selection of Lhe ilrst threshold (threshold 1), as used at step 704, and the second threshold (threshold 2) as used in step 7] 0 may be selected to improve the quality of the groupings of the rows and to minimize the number of.ngrorped rows Threshold 1 may be lowered to minimlze the number of ngroupecl rows, and threshold 2 may be increased to improve the quality of the grouping Since selection of these two thresholds are interdependent, Lhe value selected for one varies with the other in an embodiment. It should he noted that the selection of threshold 3 may vary with each embodiment and may be characterized as being data-dependent. For example, selection of threshold 3 may be made depending on the scanning resolution, i.e. how many scans are acquired across a chromatographic peak.

At step 716, a determination is made as to whether all the columns m row "i" have been processed. If not, control proceeds to step 718 where j is increased by l and control proceeds to step 710 to examine the next element in the current row. If all columns in row "i" have been processed, control proceeds to step 702 where the next row "i" is determined.

[t should be noted that the first threshold described above in COnneCtiOn with step 704 may at'fcct the number of rows of the correlation matrix which are not included in a group.

The ungrouped rows may include, for example, noise, or individual peaks, so that raising the cutoff threshold l reduces the number of grouped rows and removes noise in the dataset prior to correlation Using the example eml1odiTnent of chsterhg or grouping described in connection with 1ig'.ue 12, the first. and second i:hrcsholds in the grouping or chstering] 5 processing affect the mrnler of ungroupefl rows Threshold 1 and threshold 2 troth vary between O and 1. rl'hc first threshold, threshold 1, is the threshold for choosing a row as having valid data, and the second threshold, threshold2, is the threshold for grouping one TOW with another. Threshold 3 is the maximum separation (jet scans or seconds) allowedbetween a row's chromatographic peak and the seed row's chromatographic peak.

What will now be described is a simplified example in which the method steps described herein are performed utilizing an initial data set in matrix form. In the folkJwing example, it is assumed that there is no filtering perl'ormed in comlection with steps 404 and 408. Additionally, note that the data set used herein is not a typical data set but a small sample matrix selected for illustrative purposes of utilizing the techniques described herein.

The correlation step 406 and grouping or clustering step 410 are now perfonned using a data matrix 13 (8xX) Each row represents a mass chromatogram and each column represents a scan or time point. 13=

0 10.798 79.788 10 798 0 0 0 0 0 0 0 4.3821 99.736 4 3821 0 0 0 0 0 0 0 199.47 0 0 0 32.395 239 37 32.395 0 0 0 0 0 0 1().7')8 79.788 10 79X () O O 0 0 0 0 0 398.94 0 0 0 398.94 0 0 0 0 0 0 0 21.596 159.58 21.596 0 0 0 0 correlation matrix (8x8), (I, Is created as a result ol shelf 406 processing The resulting matrix C is: 1 -0.19738 -0.18584 1.()82468 -() 18584 -().027494 -0 1')738 1 -0.1()636 -0.1'3738 0 0076672 -0 10636 -0 15713 -O 18584 -).10636 1 -0 18584 -0.18584 1 -0.14286 0.18584 1 -0.19738 -0.18584 1 0 082468 -0.18584 -0.027494 1 0.082468 0.0076672 -0.185840.082468 1 -0.18584 -0.18584 0.082468 -0.18584 -0.10636 1-0 18584 -0.18584 1 -0.14286 0.18584 -0.027494 -0.15713 -0.14286-0.027494 -0.18584 -0.14286 1 0 027494 1 -0.19738 -0.18584 1 0.082468 -0.18584 -0.027494 The grouping or clustering steps of flowchart 7()0 may be performed to group particular rows of the correlation matrix C together. A group index vector (group) having a number of entries equal to the number ot rows in the correlation matrix may be used to indicate which rows in the correlation matrix belong to which groups Phis indication may be made by having a group number m each entry and the n-th entry of the group index vector identifies the group number ol the itch row of the correlation matrix.

Continumg with the foregoing example, the associated group vector Is group-- I 0 2 1 0 2 0 1 I o illustrate this further, the correlation matrix ('I may be reordered according to the labels in the associated group vector, in order demonstrate the nature of the grouping algorithm: C] = 1 1 1 -0.18584 -0.18584 -0.027494 -0.19738 0.082468 1 1 1 -0.18584 -0.18584 -0.027494 -0.19738 0 082468 5] 1 1 -0.] 8584 -0.18584 -0.027494 -0.] 9738 0.082468 -0.18584 -0.185X4 -0.18584] 1 -0.14286 -0.10636 0.18584 -0.18584 -0.18584 -0 18584 1 1 -0.14286 -0.10636 100.]8584 -() 027494 -0.027494 -0.027494 -0.14286 -0.14286 1 -0 15713 0.]8584 -0]9738 -0.]9738 -0 19738 -0.1()63fi -().10636 -0.15713 1 ().()()76672 150082468 ()()82468 00824(iS -() 18584 -().IX.SX4 -() IX.SX4 ()()()7(i672 Referring now to Figures 13-17, shown are example graphical displays of a data set at different points in processing when performing the method steps of Figure 10. Figure 13 shows a sample input data set 1000 that nay be gcncratcd as a result of step 402 processing.

After filtering at step 4()4, the original data set may be represented as in example display 11 ()() of Figure 14 After the correlation processing step 406, the correlation matrix may be graphically represented as 1200 in Figure 15 After identifying groups of clusters lay performing the method steps of flowchart 700 of Figure 11, the resulting groupings nay be graphically ilhstrated by reordering the correlation matrix as In 1300 of} figure 16. the filtered data may be grouped according to the group vector which results from performing the steps of flowchart 700 The example display 1400 of Figure 17 represents the reordered m/z rows such that m/z rows m the same group are adjacent. After selecting relevant scan(s) for each group, the corresponding intensities for the selected scans may be obtained from the filtered data set to produce a resulting spectra. In one embodiment as described herein, the scans may be selected by finding the scan or time at which each group maximizes the correlation vahe by adding the rows of the data matrix for each group and selecting the scan with the maximum IO intensity value.

The foregoing processing techniques described herein, for example, in connection with flowchart 400, may not be used in instances where there arc two or more molecules that clutc at the same t.imc anal also have the same ehton prol'ile. In this instance, the foregoing processing steps are not able to identl'y the dil'ferent peptides and properly pair parent (I J spectra) with l'ragments (It spectra), and another processing technique may be used, for example, as described in Attorney I locket No. 100205151 (2003309-0034), AGS-OO IOI U.S. Patent Application No. 10/388,088, filed March 13, 2003, entitled " Methods and Devices for [dentifying 13iopolymers Using Mass Spectroscopy", hereinafter referred to as 2() "the 'Thompson and Fischer clisclosurc". The processing steps ofI1, ompson and lischer may be performed on the results produced by processing steps described herein to resolve the parent-fragment pairings in instances where two or more molecules elutc at the same time.

The Thompson and Fischer disclosure describes a method for gathering structural information for biopolymers in a sample by running the mass spectrometer in the alternating scan mode, as described elsewhere herein, with alternating U and I' spectra Alternating scan mode provides for taking a first spectrum (U spectrum) at normal energy levels, such that fragmentation is not induced, and then a next second scan is taken at high fragmentation energy levels (F spectrum) where energy is injected by increased voltage differential between components of the ionization source, frequency stimulation, or some other technique producing a sequence of alternating spectra that can be deconvolved or decomposed to associate the appropriate fragment Ions from the Ii spectrum with the proper parent in the U spectrum.

When using an input data set that includes alternating scan mode data, the tcchmque described herein may be a preprocessing step performed prior to the method described m the Thompson and Fischer disclosure to associate the proper parent with the fragments (pairings of U and F spectra). (charge assignment, isotope clustering, de novo sequencing, data base searching, and the like may subsequently be performed.

A U spectrum includes peaks that correspond to some and preferably all of the polypeptides in the sample when these polypeptides arc Fragmented A 1) spectrum may he J 5 obtained by detecting the polypeptides m the sample without exposing them to a fragmentation mechanism It is to be understood that a U spectrum may, in certain embodiments, include peaks that represent fragments of these polypeptides, e g, fragments that were Inadvertently created as a consequence of the mechanism used to ionize and/or detect the polypeptides in the spectrometer.

An F spectrum includes peaks that correspond to a collection of fragments of some anti preferably all of the polypeptides in the sample An 17 spectrum may be obtained by detecting the polypeptides in the sample after these have been exposed to one or more fragmentation mechanisms It is to be understood that an F spectrum may, in certain embodiments, inchde peaks that represent unfragmented polypeptides, e.g. polypeptides that survive exposure to the fragmentation mechanism It will be appreciated that such situations are most likely to occur when the po] ypeptides are exposed to relatively low fragmentation energies 'I'he processing techniques described herein may also be performed using input data sets with multimodal chromoatograms characterized as Ions or sets of ions of the same m/z value but having different chemical compositions. Graphically, a multimodal curve has multiple peaks, for example, such as if curve 3 of Figure 9 had multiple peaks rather than the single peak as shown In the display 350. An additional step to the flowchart 10 may be used to detect multimode] curves, for example, prior to step 406 where correlation is performed In the event the multimodal curves are determined, additional processing Is performed on the input data sets In particular, additional processing Is perfonned prior to performing step 406 and as part of constructing the resultant spectra at step 414 This additional and modified p.cssig is clcscilc1 flill<'wi'p, 1'?,'plS

IS

Multimodal peaks may be detected by using a peak finding t:eclmique which determines that a particular row of tle original Input data set has multiple peaks in a single curve Although any one of a variety of different techniques may be used, one embodiment detects peaks by first filtering a row so that a baseline is removed causing peaks to be 2() separated by zero vahes An end of a peak may be determined by finding the scan at which the first derivative indicating slope of a lme, Is negative If multimodal curves are determined in a particular row of the original data set, prior l:o performing correlation step 406, the two curves may be separated by, for example, splitting the row of original data into multiple rows, one for each additional peak. 'lithe row is split after each peak in the chromatogram. 'lathe remaining Intros m each row may be zero filled. Alternatively, an embodiment may utilize other techniques, such as interpolation and curve fitting techniques, to fill in the remaining entries For example, consider a row of data in the original data matrix as described herein as follows: entry # 1 2 3 4 5 6 7 8 9 10 n 0 3.0 2.4 10.0 3.0 1.0 4 0 20 0 2.2 3.4 and that the peak finding technique determines that there a multiple peaks corresponding to elements 4 and 8 above with values, respectively, of 10.0 and 20.0. One example embodiment may, in this instance, split the foregoing row of data into two rows with a first row including elements I through 6, and a second row including elements 6 - n. The remaining cleTnents in the first and second rows may be zero filled or otherwise determined In accordance with particular techniques, such as curve fitting and interpolation, to correct the curves and provide missing data elements. Different curve fitting techniques are well known and described, for example, in the text by C. Daniel and r'.S. Wood, "Fitting l.quations to Dala" Jolm W'ley and Sons, New Stork, 1')80 An embodiment may include a multimodal detection and correction technique that may be implemented using hardware and/or software. Ibis row splitting allows a single chroTnatogram to be a member of multiple groups.

2() Another cTnbodiTnent may include the use of image processing algorithms, such as the watershed algorithm, to perform peak finding in the tinTe and m/z dimensions simultaneously.

This approach would avoid the need to perfonn the aforementioned technique of peak splitting by performing the peak finding. Additionally, it would serve to partition the dataset into peaks, thereby reducing the size of the correlation matrix. T his algorithm as well as other image processing techniques are described m K R. Castleman, "Digital Image Processing" Prentce-Hall Inc., New Jersey 1996.

In connection with step 414 processing to produce a resultant spectra, the original data set is again utilized.

In particular, as described elsewhere herein, the appropriate columns of intensities for the selected scans are obtained from the original data set. With multimodal data, it should be noted that an m/z range may appear in more than one group.

An embodiment may utihze any one of different types of mass spectra that may be produced, for example, by a tme-of-flight (TOF) mass spectrometer. An example embodiment may use include a step following step 402 in which input data sets are converted to a more compact form prior to be used with the foregoing processing steps. For example, a I OF data set may lie converted to lee utihzed with ibe foregoing techniques. The TOF input data set may be a 2-dimensonal matrix With the Y-axis ndicatmg the time of flight correlating directly to the m/z values and the elation time on the x-axis. Each column of the TOF data is a scan of the mass spectrum data. This matrix may be converted into a sparser fomm to minimize storage. The compaction technique used on the matrix may vary in accordance with the functionality and particular components included in each embodiment One example embodiment utilizes a MATLAB function to compress the matrix into a sparse matrix format. Any needed subsequent conversions may be performed by MATLAB. An embodiment may optionally use other formats depending on memory constraints and other characteristics of an embodiment.

An embodiment may utilize filtering techniques to reduce noise and eliminate data associated with known contaminants For example, particular correlation values of a known contaminant withm a certain m/z range may be eliminated at step 408 Consider, for example, that a known detergent contaminant may be present. The contaminant presence may be determined by manually examining a contour plot and visually locating a constant horizontal band present at all elusion times. Input data sets may be examined to automatically test for known contaminants and accordingly remove the bands of data. It should be noted that an example embodiment may provides for "noise" to be filtered that is highly correlated, such as a known contaminant, and/or weakly correlated, such as interference.

It should be noted that the techniques described herein may be used for performing a quantitative analysis rather than for Identification processing, for example, such as identifying matching F and U spectra. This may affect the previously described processing steps. When perfonning a quantitative analysis using the foregoing techniques, points of I interest selected, as at step 412, may include those sampled frequently across each group, rather than determining a single maximum as described herein. As described elsewhere herein, step 414 processing produces a single spectra for each ion with contaminants and other covarying spectra removed. For quantitative analysis using the foregoing techniques, a spectrum is produced for each cluster or group. For quantitation, the peak areas are integrated for the group chromatograms or rows. This provides a group peak area that may be used for relative quantitation with other groups in the data set. For quantitation, each cluster or group using the foregoing techniques represents a range of m/z values and elusion time that contains related signal.

The foregoing provides techniques utilizing the fact that certain groupings tend to covary. Parent and related ion fragments tend to covary and exhibit similar coelution profiles Input data including only U spectra, when processed by the techniques described herein, may be used to group charge states and Isotopes of single peptides since these charge states and isotopes covary by coeluting at the same time Input data including only F spectra may be used to group charge state, isotopes and fragments that coelute at the same time. The foregoing may also be used as a preprocessing step in connection with the Thompson and Fischer disclosure and other processing techniques to identify U and related F spectra when two parent or U spectra within a group have the same elusion profile and coelute at the same time. Such other techniques may include, for cxamplc, identification algorithms, such as SEQIIEST, MASCOT, MSFIT, and the like. these techniques are known in the art. For example, SEQUEST is described in: Eng, J. K.; Mc(:ormack, A. L.; Yates J. R. III. J Am Soc Mass Speciorm. 1994, 5, 976-989, MASCOT is described in: Perkins, D. N.; Pappin, D. J. C., Creasy, D. M.; Cottrell, J. S Electrophores.s I 999, 20, 3551 - 3567, and MSFI I is described in: (lauser K. R., Baker P R. and Burlingame L., Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Analytical Chemistry, Vol. 71, 14, 2871- (1999).

Use of the Thompson and Fischer disclosure and/or other teclmique may be used to distinguish between two unrelated components (not isotopes, charge states or fiagments) that coelute and exactly covary since the techniques described herein will not be able to distinguish between two such unrelated compounds. Different techniques may be used to determine the existence of such a condition indicating a need to invoke alternative techniques to assign these parents to their corresponding fragments. An embodiment may test extracted U spectra for the presence of multiple parents which the foregoing techniques cannot distinguish between as follows. Deisotoping and charge deconvolution may be performed on the spectrum

Method and program for identifying ions from chromatographic mass spectral data sets

Claims of **GB2403342** Translate this text

CLAIMS

1. A method for identifying related ions in an input data set produced by analyzing a sample comprising: correlating each row of data in an Input data set with every other row of data in said input data set producing a correlation matrix, each row representing intensities over hme for a particular mass to charge (m/z) range, each element of said correlation matrix including a correlation value and having associated row and column identifiers dentifymg which rows in said input data set are associated with said correlation value; clustering said correlation matrix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each group representing covarying chron1atograms; selecting at least one time period of Interest for each group; and producing a resultant spectrum for each group by sampling chromatograms included in each of said groups at each of said at least one time period of interest of uslng a form of said input data set.

2. The method of Claim 1, further comprising filtering said input data set prior to performing said correlation.

3. The method of Claim 1 or 2; wherein said input data set includes only one of: unfragmenteA spectrum, fragmented spectrum, and alternating unfragmented and fragmented spectrum.

4. The method of Claim 3, wherein said input data set includes only alternating unfragmented and fragmented spectra, and the method further comprising: forming a combined spectrum including an unfragmented spectrum and related fragmented spectrum; performing said correlating, said clustering, said selecting and said producing using said combined spectrum; determining m/z values in said combined spectrum; sampling said unfragmcrted spectra at said rn/z values in said comli,led spectrum at a scan maxmunl identified for said combined spectrum; and detenninmg that a sampled mlz value in said combined spectra Is associated with a parent if there is an intensity at said sampled m/z value, and determining that said sampled m/z value in said combined spectra is associated with a fragment in an absence of a signal at said sampled m/z value.

5. The method of Claim 3, wherein said input data set includes only alternating unfragmented and fragmented spectra, and the method further comprising: perfonning said correlating, said clustering, said selecting and said producing using each of said unfragmented spectrum and said fragmented spectrum separately; and matching each parent of sald unfragmented spectrum to related fragments m said fragmented spectrum by determining which of said related fragments covary with said parent.

6. The method of any preceding Claim, wherein said clustering further comprises: determining a first row of said correlation matrix including an element having a maximum correlation value of all correlation values in the correlation matrlx being considered as candidates to be grouped; determining a time scan assocatcd with said first row; for each element of said first row corresponding to a unique pairing of a row "i" and column ";", determining if a correlation value is greater than a predetermined value and determining if a scan number at which row "j" maximizes is within a threshold number of scans of said time scan associated with said first row; and If said each element is greater than said predetermined value and If a scan number at which row "j" maximizes is within a threshold number of scans of said time scan associated with said first row, adding a row of said input data set to a current group wherein the row added has a row number equal to that of a column index "j" associated with said each element, and excluding the row added from further consideration as one of said candidates for grouping.

7. The method of Claim 6, further composing: performing said determining a first row if a correlation value is greater than a predetermined value and if a scan number at which row "j" maximizes Is within a threshold number of scans of said time scan associated with said first row for each element of the first row having an associated column index greater than an index associated with said first row.

8. The method of Claim 7, further comprising: stopping formation of groups by said clustering when said maximum correlation value Is less than a predetermined value.

9. The method of Claim 8, further comprising: forming a new group with a selection of a subsequent row including an element having a maximum correlation value of all correlation values in the correlation matrix bemg considered as candidates to be grouped.

10. The method of Claim 3, wherein said input data set includes only alternating unfragmented and fragmented spectra, said input data set includes at least two components elating at a same time and having a same elusion profile, and the method further comprising: combining adjacent fragmented and unfragmented spectra resulting in a new combined data set including half the number of spectra in comparison to a total of spectra of said fragmented and unfragmented spectra; producing a first resulting spectrum and a second resulting spectrum, said first resulting spectrum corresponding to said unfragmented spectrum at a selected point in time and said second resulting spectrum corresponding to said fragmented spectrum at said selected point In time; and performing processing to Identify which of said at least two components Is a parent associated with at least one fragment included in said fragmented spectrum.

11. The method of Claim 3, wherein said input data set includes only unfragmented spectrum, and said at least one group formed by said clustering Identifies charge states and Isotopes of a single component that coelute at a same time.

12. The method of Claim 3, wherein said input data set includes only fragmented spectrum, and said at least one group formed by said clustering identifies charge states, isotopes, and fragments of a single component that coelute at a same time.

13. The method of any preceding Claim, wherein said selecting time periods of interest includes: summing intensities of extracted chromatogrrams for each group at each scan point; and determining a maximum intensity for each group at a particular scan point; and wherein said producing a resultant spectrum includes: sampling extracted chromatograms of each group at said particular scan point.

IO 14. The method of any preceding Claim, wherein said input data set is produced using a mass spectrometer analyzing the sample.

15. The method of any preceding Claim, wherein said input data set includes at least one multimodal peak of an extracted ion chromatogram, a number of peaks in said multimodal peak being represented as "n", and the method further comprising: deterrninmg at least one split point in said multimodal peak to dlvlde sald multimodal peak Into portions; apportioning a first row of said input data set corresponding to said multimodal peak into row portions in accordance with said at least one split point; creating an additional "n-l" rows of data included in said input data set, each of said additional rows including a different one of said row portions; removing from sald first row all row portions included in said additional rows; and filling remaining elements of each of said additional rows and said first row.

16. The method of any preceding Claim, further comprising: filtering said input data set producing a filtered data set, and wherein said form of said input data set is said filtered data set.

17. The method of Claim 10, wherein at least two ions are two parent ions co elating at a same time having a same elusion profile and covary, and the method further comprlsmg: performing other processing steps to associate each of said two parent lons witl corresponding fragment foes.

18. The method of Claim 10, wherein said at least two components are parent peptides that coelute at a same time and exhibit similar elusion profiles, and the method further comprising: determining that additional processing is needed to match each of said at least two parent peptdes with associated child fragments; and performing said additional processing.

19. The method of Claim 10, wherein said at least two components are peptides.

20. A method for quantifying at least one ion in an input data set produced by analyzing a sample comprising: correlating each row of data in an input data set with every other row of data in said input data set producing a correlation matrix, each row representing intensities over time for a particular mass to charge (m/z) range, each element of said correlation matrix including a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said correlation value; clustering said correlation matrix identifying at least one group and at least one row of

said correlation matrix as being m said at least one group, each group representing chemically related components exhibiting correlated chromatographic behavior; selecting at least one time period of interest for each group; and producing a resultant spectrum for each group by sarmplmg chromatograms included in each of said groups at each of said at least one time period of interest of using a form of said input data set.

21. A computer program product for identifying related ions in an input data set produced by analyzing a sample comprising: machine executable code that correlates each row of data in an input data set with every other row of data in said input data set producing a correlation matrix, each row representing intensities over time for a particular mass to charge (m/z) range, each element of said correlation matrix including a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said correlation value; machine executable code that clusters said correlation matrix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each group representing covarying chromatograms; machine executable code that selects at least one time period of interest for each group; and machine executable code that produces a resultant spectrum for each group by sampling chromatograms included in each of said groups at each of said at least one time period of interest of using a form of said input data set.

22. The computer program product of Claim 21, further comprising: machine executable code that filters said input data set prior to performing said correlation.

23. The computer program product of Claim 21 or 22, wherein said input data set includes only one of: unf.ragmcnted spectrum, fragmented SpecLram,, and alternating unfragmented and fragmented spectrum.

24. The computer program product of Claim 23, wherein said input data set includes only alternating unfragmented and fragmented spectra, and the computer program product further comprising: machine executable code that forms a combined spectrum including an unfragmented spectrum and related fragmented spectrum and wherein said machine executable code that correlates, clusters, selects and produces uses said combined spectrum; machine executable code that determines m/z values in said combined spectrum; machine executable code that samples said uTlfragment:ed spectra at said mlz values In said combined spectrum at a. scan maximum Identified for said combined spectrum; and machine executable code that determines that a sampled m/z value in said combined spectra is associated with a parent if there is an intensity at said sampled mlz value, and determines that said sampled m/z value in said combined spectra is associated with a fragment in an absence of a signal at said sampled m/z value.

25. The computer program product of Claim 23, wherein said input data set includes only alternating unfragmented and fragmented spectra, said machine executable code that correlates, clusters, selects and produces uses each of said unfragmented spectrum and said fragmented spectrum separately; and the computer program product further comprising: machine executable code that matches each parent of said unfragmented spectrum to related fragments in said fragmented spectrum by determining which of said related fragments covary with said parent.

26. The computer program product of any of Claims 21 to 25, wherein said clustering further comprises: I O machine executable code that detennines a first row of said correlation matrix including an element having a maximum correlation value of all correlation values In the correlation matrix being considered as candidates to be grouped; machme executable code that determines a time scan associated with said first row; machine executable code that, for each element of said first row corresponding to a unique pairing of a row "i" and column "j", determines if a correlation value is greater than a predetermined value and determining if a scan number at which row "I" maximizes is within a threshold number of scans of said time scan associated with said first row; and machine executable code that, if said each element Is greater than said predetermined value and if a scan number at which row "j" maximizes is within a threshold number of scans of said time scan associated with said first row, adds a row of said input data set to a current group wherein the row added has a row number equal to that of a column index "j" associated with said each element, and excluding the row added from farther consideration as one of said candidates for grouping.

27. The computer program product of Claim 26, further comprising: machine executable code that determines said first row if a correlation value is greater than a predetermined value and If a scan number at which row "j" maximizes is within a threshold number of scans of said time scan associated with said first row for each element of the first row having an associated column index greater than an index associated with said first row.

28. The computer program product of Claim 27, further comprising.

machine executable code that stops formation of groups by said clustering when said rmaxmum correlation value is less than a predetermined value.

29. The computer program product of Claim 28, further comprising: machine executable code that forms a new group with a selection of a subsequent row including an element havmg a maximum correlation value of all correlation values m the correlation matrix being considered as candidates to be grouped.

30. The computer program product of Claim 23, wherein said input data set includes only alternating unfragmented and fragmented spectra, said input data set includes at least two components eluting at a same time and having a same elusion profile, and the computer program product further comprising: machine executable code that combines adjacent fragmented and unfragmented spectra resulting in a new combined data set including half the number of spectra in comparison to a total of spectra of said fragmented and unfragmented spectra; machine executable code that produces a first resulting spectrum and a second resulting spectrum, said first resulting spectrum corresponding to said unfragmented I O spectrum at a selected point in time and said second resulting spectrum corresponding to said fragmented spectrum at said selected point In time; and machine executable code that performs processing to Identify which of said at least two components Is a parent associated with at least one fragment Included in said fragmented spectrum.

31. The computer program product of Claim 23, wherein said input data set Includes only unfragmented spectrum, and said at least one group formed by said clustering identifies charge states and isotopes of a single component that coelute at a same time.

32. The computer program product of Claim 23, wherein said input data set includes only fragmented spectrum, and sald at least one group formed by said clustering identifies charge states, isotopes, and fragments of a single component that coelute at a same time.

33. The computer program product of any of Claims 21 to 32, wherein said selecting time perlods of interest includes: machine executable code that sums mtensites of extracted chromatograms for each group at each seen point; and machine executable code that determines a maximum intensity for each group at a particular scan point; and wherein said producing a resultant spectrum includes: machine executable code that samples extracted chromatograms of each group at said particular seen pomt.

34. The computer program product of any of Claims 21 to' 33, wherein said input data set Is produced using a mass spectrometer analyzing the sample.

35. The computer program pro:luet of any of Claims 21 to 34, wherein said Input data set includes at least one multimodal peak, wherein said input data set includes at least one multimodal peak of an extracted ion ehromatogram, a number of peaks in said multimodal peak being represented as "n", and the computer program further comprising: machine executable code that determines at least one split point in said multimodal peak to divide said multimodal peak into portions; machine executable code that apportions a first row of said input data set corresponding to said multimodal peak into row portions in accordance with said at least one split point; machine executable code that creates an additional "n-1" rows of data included In said input data set, each of said additional rows including a different one of said row portions; machine executable code that removes from said first row all row portions included in said additional rows; and machine executable code that fills remaining elements of each of said additional rows and said first row.

36. The computer program product of any of Claims 21 to 35, further comprising machine executable code that filters said Input data set producing a filtered data set, and wherein said form of said input data set is said filtered data set.

IS

37. IPhe computer program product as claimed in claim 30, wherein at least two components are two parent Ions co-eluting at a same time having a same elusion profile and covary, and the computer program product further comprising: I' machine executable code that performs other processing steps to associate each of said two parent ions with corresponding fragment lons.

38. The computer program product of Claim 30, wherein said at least two components are parent peptides

that coelute at a same time and exhibit similar elusion profiles, and the computer program product further comprising: machine executable code that determines that additional processing is needed to match each of said at least two parent peptides with associated child fragments; and machine executable code that performs said additional processing.

39. The computer program product of Claim 29, wherein said at least two components are peptides.

40. A computer program product for quantifying at least one ion in an input data set produced by analyzing a sample comprising: machine executable code that correlates each row of data in an input data set with every other row of data in said input data set producing a correlation matrix, each row representing intensities over time for a particular mass to charge (m/z) range, each element of said cor elation matrix including a correlation value and having associated row and column identifiers identifying which rows in said input data set are associated with said cor elation value; machine executable code that clusters said cor elation mat ix identifying at least one group and at least one row of said correlation matrix as being in said at least one group, each g oup representing chemically related components exhibiting correlated chromatog aphic behavior; machine executable code that selects at least one time period of interest for each group; and machine executable code that produces a resultant spectrum for each group by sampling chromatograms included m each of said groups at each of said at least one time period of interest of using a form of said input data set.

41. A method for identifying related ions in an input data set produced by analyzing a sample substantially as hereinhefore described with reference to and as illustrated in the accompanying drawings.

42. A method for quantifying at least one ion in an input data set produced by analyzing a sample substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.

43. A computer program product substantially as hereinbefore described with reference to and as illustrated in the accompanying drawings.

Data supplied from the ***esp@cenet*** database - Worldwide

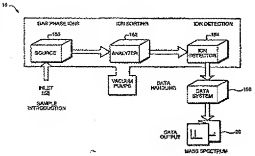

FIG. 4

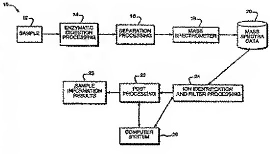

FIG. 1

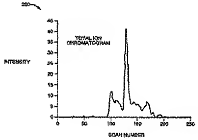

FIG. 5

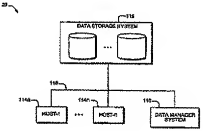

FIG. 2

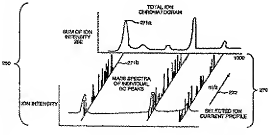

FIG. 6

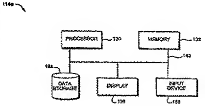

FIG. 3

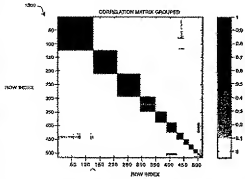

FIG. 16

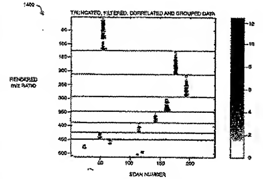

FIG. 17

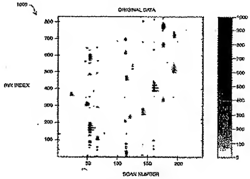

FIG. 13

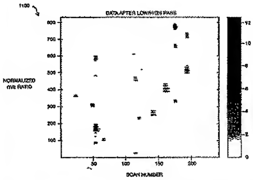

FIG. 14

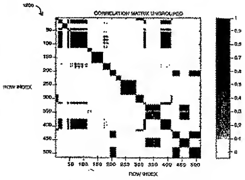

FIG. 15